# 1 Introduction

#### Mining association rules using ARS, from the SIPINA distribution.

<u>SIPINA</u> is known for its decision tree induction algorithms. In fact, the distribution includes two other tools that are little known to the public: REGRESS, which is specialized in multiple linear regression, we described it in one of our tutorials<sup>1</sup>; and an association rules extraction tool, called simply Association Rule Software (**ARS**).

At that time (1997-1998), my idea was to build a set of independent executables organized around the same data management grid. Various tools for factor analysis and clustering, in addition to the tools which are included in the current distribution, were developed. But afterward, I realized that a functioning based on visual programming using streams of components was more advantageous in many respects, in terms of ease of use for the users, but also in terms of software evolution. TANAGRA was developed with this in mind.

This does not mean that the tools included in the SIPINA distribution are not interesting. In this tutorial, I describe the use of the ARS tool. Its interactivity with Excel spreadsheet is its main advantage. We launch the software from Excel using the "sipina.xla" add-in<sup>2</sup>. We can easily retrieve the rules in the spreadsheet. Then, we can explore them (the rules) using the Excel data handling capabilities. The ability to filter and sort rules according to different criteria is a great help in detecting interesting rules. This is a very important aspect because the profusion of rules can quickly confuse the data miner.

# 2 Dataset

The "market\_basket.xlsx" data file describes the contents of n = 1361 shopping carts (transactions). We have a library of p = 303 products (items).

In average, each cart contains 9,5 products (min = 0, max = 303). And the products are purchased 42,7 times in average (min = 7, max = 167). The 5 most popular products are: "Eggs", "White bread", "2pct milk", "Potato chips" and "98pct fat free hamburger". The least popular ones are: "Celery", "Oats and Nuts Cereals", "Chicken legs", "Nasal spray" and "Daily Newspaper".

# 3 Mining association rule with ARS

### 3.1 Data importation

We load the "market\_basket.xlsx" data file into Excel (we must use Excel 2007 or later version because the number of items [p = 303] exceeds the limitation of 256 columns of Excel 2003 and earlier versions). <u>Note:</u> If you do not have the Excel 2007 version, the best solution is to import the "market\_basket.txt" data file (tab separated text format) via the "FILE / OPEN / TEXT FILE FORMAT (\*.txt)" menu<sup>3</sup>. This file is included in the archive that comes with this document.

<sup>&</sup>lt;sup>1</sup> http://data-mining-tutorials.blogspot.fr/2011/08/regress-into-sipina-package.html

<sup>&</sup>lt;sup>2</sup> http://data-mining-tutorials.blogspot.fr/2016/06/sipina-add-in-for-excel-2007-and-2010.html

<sup>&</sup>lt;sup>3</sup> Sipina – Supported file format - <u>http://data-mining-tutorials.blogspot.fr/2009/11/sipina-supported-file-format.html</u>

After loading the data file into Excel, we select the data range and click the menu SIPINA / ASSOCIATION RULE. A setting dialog box appears. We check the coordinates of the data range and we click the OK button.

C		market_basket.	xlsx - Microsoft Excel		
r <sup>c</sup>	Accueil Insertion Mise er	n pag Formules Données	Révision Affichage	Développeu Complémen	1 🞯 🗕 📼 🗙
	Sipina 🔻				
	Execute Sipina				
	Association Rule				
Cor	Regression				
	About	<i>f</i> ∗ Hair_Conditi	oner		*
	A	Launch Data Mining Too	d		
1	Hair_Conditioner Lemor			tation front and be	
2	0	Lataset range (inc	luding the name of the at	tributes first row):	
3	0	9491.90291302	·		-
5	0		c	OK Cancel	
6	0				_
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	1
10	0	0	0	0	0
11	0	0	0	0	0
12	\mapsto भ dataset 🕅			0	▶
Prêt	t Moyenne : 0.031366472	Nb (non vides) : 412686	Somme : 12935	100 % 🕞	

The software is automatically launched and the dataset is imported. We check that we have n = 1361 rows and p = 303 columns.

File Edit Data Statisti	cs Analysis	Window	Help						- 8
🖻 🔒									
Selected attributes		Hair_Con	ditic Lemons	Standard	_co Frozen_C	hic 98pct_Fa	t_F Sugar_Co	oki Onions	Deli
	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	8	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Learning Parameters	11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	13	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00
	14	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00
	15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Selected examples	18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	•	-					1		
	Editing	NEW EDM				Attr	ibutes : 303	Examples	:1361

## 3.2 Selecting the items

To select the items to include into the analysis, we click the ANALYSIS / SELECT ATTRIBUTES menu. We select all the columns.

Association rule software - [Le	arning set editor]	
👶 File Edit Data Statistics	Analysis Window H	Help _ B ×
D 😅 🖬	Select attributes	
Selected attributes	Select examples	Standard_co Frozen_Chic/98pct_Fat_F Sugar_Cooki Onions Deli 1 ^
	Start analysis	Select attributes for an Usis
Learning Parameters	3         0.00           4         0.00           5         0.00           6         0.00           7         0.00           8         0.00           9         0.00           10         0.00           11         0.00	Hair_Conditioner Lemons Standard_coffee Frozen_Chicken_Wings 98pct_Fat_Free_Hamburger Sugar_Cookies Onions Deli_Ham Disthwasher_Detergent Beets 40_Watt_Lightbulb toe_Cream Cottane_Cheese
Selected examples	12         0.00           13         0.00           14         0.00           15         0.00           16         1.00           17         0.00           18         0.00	Plain_English_Muffins       Strawberry_Soda       Vanilla_Ice_Cream       Potato_Chips       Strawberry_Yoguit       Diet_Soda       D_Cell_Batteries
	editing NEW.FDM	WARNING: Numeric attributes are treated as binary ones. Ok Cancel

Because our variables are numerical, ARS takes them as dummy variables: o, absence of the item in the shopping cart; 1 (it can be any value > o), presence. When we process categorical variables, the tool automatically performs a dummy coding before starting processing.

## 3.3 Settings for the rule extraction process

Association rule software - [Lea	arning set e	ditor]							) <b>X</b>	<u> </u>
🔮 File Edit Data Statistics	Analysis	Window	Help						- 8	×
🗅 🚔 🔒	Sele	ct attribute	s							
Selected attributes	Sele	ct example	s	Standar	d_or Frozen_0	Chic 98pct_Fa	It_F Sugar_Co	ooki Onions	Deli_i	-
Hair_Conditioner	Star	tanahesis		0.00	0.00	0.00	0.00	0.00	0.00	
Standard coffee	Star	analysis		Accocia	tion Rule Darar	materr	0.00	X	0.00	
Frozen_Chicken_Wings	3	0.00	0.00	Associa	Nule Falar	neters			0.00	
98pct_Fat_Free_Hamburger	4	0.00	0.00	Learni	ng Measures				0.00	
Diplops	5	0.00	0.00						0.00	
Deli Ham	6	0.00	0.00		Support :		0.0	115 🛑 🗌	0.00	
Dishwasher_Detergent	7	0.00	0.00							
Beets 40. ) (att. Liabthulb	8	0.00	0.00		Confidence	:	0.	.85	0.00	
lce Cream	9	0.00	0.00						0.00	
Cottage Cheese	10	0.00	0.00		Max rule ler	ngth :	4	2	0.00	
Learning Parameters	11	0.00	0.00		Max conser	uent length :	1		0.00	
	12	0.00	0.00			quorit iorigan.			0.00	
	13	0.00	0.00						0.00	
	14	0.00	0.00			0.	Consel		0.00	
	15	0.00	0.00				Lancel		0.00	
	16	1.00	0.00						0.00	
	17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Selected examples	18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Ŧ
	•								•	
<u> </u>	Editing	NEW.FDM				Att	ributes : 303	Examples :	1361	1.
										///

To launch the association rule learning, we click the ANALYSIS / START ANALYSIS menu. A setting dialog box appears: we set the minimum support to 0.015 (we accept the rules which occurs at least 1361 x 0.015  $\approx$  20 times into the database); the minimum confidence is 0.85; the maximum cardinal of the rule is 4 items; we have at most 1 item into the consequent of the rule.

Analysis Rule manager	ment	Wir	ndow Help								- 8	>
🖻 🚔 🔚												
Selected attributes	[	Ba I										
air_Conditioner		Id	Antecedent	Consequent	Length	Support	Confi	Recall	F-mea	Lift	Convi	Τ
andard_coffee		1	AA_Cell_Batteries & Eggs	2pct_Milk	3	0.0169	0.8519	0.1544	0.2614	7.7810	5.1617	Ĩ.
ozen_Chicken_Wings		2	Apple_Jelly & Wheat_Bread	2pct_Milk	3	0.0176	0.9231	0.1611	0.2743	8.4316	8.3062	
Spot_Fat_Free_Hamburger		3	Apples & Onions	2pct_Milk	3	0.0191	0.8667	0.1745	0.2905	7.9163	5.6957	
nions		4	Apples & Potato_Chips	2pct_Milk	3	0.0228	0.8611	0.2081	0.3351	7.8656	5.6363	
eli_Ham ishuusehar Dataraant		5	Bananas & Onions	2pct_Milk	3	0.0206	0.8750	0.1879	0.3094	7.9924	6.0517	
isriwasriei_Deteigent eets		6	Bananas & Wheat_Bread	2pct_Milk	3	0.0220	0.8571	0.2013	0.3261	7.8293	5.4880	
)_Watt_Lightbulb		7	Cantaloupe & Pepperoni_PizzaFrozen	2pct_Milk	3	0.0169	0.8519	0.1544	0.2614	7.7810	5.1617	
e_Liream	-	8	Cantaloupe & Wheat_Bread	2pct_Milk	3	0.0184	0.8929	0.1678	0.2825	8.1556	6.6746	
Learning Parameters		Э	Cream_Cheese & Popcorn_Salt	2pct_Milk	3	0.0176	0.8889	0.1611	0.2727	8.1193	6.4521	
in support = 0.0150		10	Eggs & flav_Fruit_Bars	2pct_Milk	3	0.0176	0.8571	0.1611	0.2712	7.8293	5.3397	
in confidence = 0.8500 av rule length = 4		11	Hair_Conditioner & Oranges	2pct_Milk	3	0.0154	0.9130	0.1409	0.2442	8.3399	7.4162	
ax cons. length = 1		12	Hair_Conditioner & Wheat_Bread	2pct_Milk	3	0.0198	0.8710	0.1812	0.3000	7.9556	5.8737	
•		13	Hot_Dogs & Shampoo	2pct_Milk	3	0.0162	0.8800	0.1477	0.2529	8.0381	6.0072	
		14	Hot_Dogs & Trash_Bags	2pct_Milk	3	0.0169	0.9583	0.1544	0.2659	8.7536	11.5693	
Selected examples		15	Pepperoni_PizzaFrozen & Popcorn_Salt	2pct_Milk	3	0.0184	0.8929	0.1678	0.2825	8.1556	6.6746	
		16	Plastic_Spoons & Toothpaste	2pct_Milk	3	0.0154	0.8750	0.1409	0.2428	7.9924	5.7847	

When we validate these settings by clicking the OK button, we obtain a new window which contains the mined rules. Each rule is characterized by its antecedent, its consequent, and a set of numerical indicators (support, confidence, lift, etc.)<sup>4</sup>.

We obtain **510 rules**. The first one (n°1) is:

IF purchase (AA\_Cell\_Batteries & Eggs) THEN purchase also (2pct\_Milk)

The **Support** of the rule is:

P(AA\_Cell\_Batteries & Eggs & 2pct\_Milk) = 0.0169

Its Confidence:

P(2pct\_Milk / AA\_Cell\_Batteries & Eggs) = 0.8519

Its Lift:

 $\frac{P(2pct_milk/AA_Cell_Batteries \& Eggs)}{P(2pct_milk)} = 7.7810$ 

Actually, nothing really distinguishes this tool from the association rule extractions components available in TANAGRA. My idea in this tutorial is to highlight in the next section the features... of Excel that allow us to examine the results (mined rules) better.

<sup>&</sup>lt;sup>4</sup> "Interestingness measures for association rules" - <u>http://data-mining-tutorials.blogspot.fr/2009/02/interestingness-</u> measures-for.html

# 4 Exploring the rules under Excel

## 4.1 Retrieving the rules

Association rules learning algorithms often generate a large number of rules, 510 for our study. To be able to examine them, we must be able to organize the rules at our convenience: filter them according to several criteria, sort them according to the measures of interestingness, etc. The Excel's **Sort & Filter** tools are really appropriate for that purpose.

First, we must copy the rules into Excel. From ARS, we click the RULE MANAGEMENT / COPY RULES TO CLIPBOARD menu.

🍓 Assoc	iation	rule	software - [Ass	ociation rule	viewer (Beta = 1.00)]									1 X	Γ
🌒 Ana	lysis	Rule	management	Window I	Help									_ & ×	1
D 🖻			Beta paramete	er for f-measu	re										
2	Selecti		Copy rules to	clibpoard	Ctrl+C										
Hair_Con	ditione		Save ruleset				Consequent	Length	Support	Confi	Recall	F-mea L	Lift C	Convi	
Standard	_coffee	-			_раценез и рудз		2pct_Milk	3	0.0169	0.8519	0.1544	0.2614 7	.7810 5.	1617 🔺	
Frozen_C	hicker	_Win ⊔ ∍m	gs burger	2 Apple_J	elly & Wheat_Bread		2pct_Milk	3	0.0176	0.9231	0.1611	0.2743 8	.4316 8.	3062 🔲	
Sugar Co	n_nice; nokies		burger	3 Apples	k Onions		2pct_Milk	3	0.0191	0.8667	0.1745	0.2905 7	.9163 5.	6957	
Onions	JOINICS			4 Apples	A Potato Chips		2pct Milk	3	0.0228	0.8611	0.2081	0.3351 7	.8656 5.	6363	
Deli_Har			· (2 · ) =	1		n	narket basket.xlsx - Mi	crosoft Excel							x
Dishwas Beets		Accue	il Insertion M	lise en page Fo	mules Données Révision	Affichage	- Développeur	Compléments							σx
40 Watt	Bààn	artir du	fichier Access				🔽 😵	Effacer		🔤 🖂 Vali	dation des don	nées 🔹 🏟 Gr	ouper 🛪 📲	2. Solveur	
Ice_Crea	🔛 À p	artir du	site Web		Propriétés	Ź		Réappliquer		→ Cor	solider	4 Di	ssocier + ==	- Jonean	
IC	Àp	artir du	À parti d	d'autres Connexion	Actualiser	d'accès A	Trier Filtrer	Avancé	Convertir Supp	orimer	llyse de scénario	s - 🗐 So	us-total		
'			Données externe	s	Connexions		Trianitre	r	103 00	Outils de don	nées	1	Plan 🕞	Analyse	
Min supp Min cont		B8	-	<i>f</i> ∗ Cantaloup	e & Pepperoni_PizzaFroze	n									×
Max rule		А		B			с	D	E	F	G	н	1	J	-
Max con	1 Id		<ul> <li>Antecedent</li> </ul>			<ul> <li>Consequ</li> </ul>	ent 🔽	Length 💌	Support 💌	Confiden	Recall 💌	F-measur	Lift 💌	Convictio	-
	2		1 AA_Cell_Batteri	es & Eggs		2pct_Mi	k	3	0.0169	0.8519	0.1544	0.2614	7.781	5.1617	
	3		2 Apple_Jelly & W	/heat_Bread		2pct_Mi	k	3	0.0176	0.9231	0.1611	0.2743	8.4316	8.3062	
	4		3 Apples & Onion	s		2pct_Mi	k	3	0.0191	0.8667	0.1745	0.2905	7.9163	5.6957	_
	5		4 Apples & Potato	_Chips		2pct_Mi	k	3	0.0228	0.8611	0.2081	0.3351	7.8656	5.6363	_
	6		5 Bananas & Onio	ns		2pct_Mi	k	3	0.0206	0.875	0.1879	0.3094	7.9924	6.0517	
	7		6 Bananas & Whe	at_Bread	-	2pct_Mi	k	3	0.022	0.8571	0.2013	0.3261	7.8293	5.488	_
	8		/ Cantaloupe & Pe	epperoni_Pizza	Frozen	2pct_Mi	K k	3	0.0169	0.8519	0.1544	0.2614	/./81	5.101/	_
	10		9 Cream Cheese	& Poncorn Salt		2pct_Mi	k	2	0.0104	0.8323	0.1078	0.2823	8 1193	6.4521	
	11		10 Eggs & flav, Eruit	t Bars		2pct_Mi	k	3	0.0176	0.8571	0.1611	0.2712	7,8293	5.3397	
	12		11 Hair Conditione	er & Oranges		2pct Mi	k.	3	0.0154	0.913	0.1409	0.2442	8.3399	7.4162	
	13		12 Hair Conditione	er & Wheat Bread		2pct Mi	k	3	0.0198	0.871	0.1812	0.3	7.9556	5.8737	
	14		13 Hot_Dogs & Sha	mpoo		2pct_Mi	k	3	0.0162	0.88	0.1477	0.2529	8.0381	6.0072	
	15		14 Hot_Dogs & Tras	h_Bags		2pct_Mi	k	3	0.0169	0.9583	0.1544	0.2659	8.7536	11.5693	
	16		15 Pepperoni_Pizz	aFrozen & Pop	corn_Salt	2pct_Mi	k	3	0.0184	0.8929	0.1678	0.2825	8.1556	6.6746	-
		N da	taset 📜 Feuil1 / 💱												
	Pret					_							100 % (=)	- V	.::

We create a new worksheet in Excel. We paste the rules. Then we activate the **FILTER** button (**Data** tab, **Sort & Filter** group)<sup>5</sup>. Arrows in columns headers enable to select various criteria for filtering the rules. The same functionalities are available in free spreadsheets such as Calc of LibreOffice<sup>6</sup> or Apache OpenOffice<sup>7</sup>.

### 4.2 Sorting the rules according to a numerical indicator

We want to sort the rules according to the LIFT criterion. We click on the arrow of the corresponding column, we ask "Sort from the largest to the smallest".

 $<sup>^{\</sup>rm 5}$  Our screenshots are based on the French version of Excel 2007.

<sup>&</sup>lt;sup>6</sup> <u>https://fr.libreoffice.org/</u>

<sup>&</sup>lt;sup>7</sup> https://www.openoffice.org

#### Tanagra Tutorial

		9-0					market_b	asket.xlsx - Micr	osoft Excel						- 0	x
		Accueil	Insertion	Mise en page	Formules	Données	Révision	Affichage	Dévelop	oeur Coi	mpléments				@ _ =	×
Don exter	nées mes *	Actualiser tout *	Connexio Propriété Modifier	ons is les liens d'accès	$ \begin{array}{c} \underline{A} \downarrow \\ \underline{Z} \downarrow \\ \underline{Z} \downarrow \\ \underline{A} \downarrow \end{array} $ $ \begin{array}{c} \underline{A} \downarrow \\ \underline{Z} \downarrow \\ \underline{A} \downarrow \end{array} $ $ \begin{array}{c} \underline{A} \downarrow \\ \underline{Trier} \end{array} $	Filtrer	Effacer Réappliquer Avancé	Convertir Supples do	orimer ublons	Validation d Consolider Analyse de s	es données 👻	<ul> <li>Groupe</li> <li>Dissocio</li> <li>Sous-to</li> </ul>	er * 📲 🏤 9 er * 📲 otal	Solveur		
			Connexio	ns	1	Frier et filtrer			Outils de	données		Plan	S Ar	alyse		
		B3	<b>-</b>	<i>f</i> ∗ Suga	r_Cookies &	Sweet_Rel	ish & White	_Bread								×
	A B C D E F G H												J	-		
1	Id 💌	Anteceden	t			-	Consequent	-	Length 💌	Support 💌	Confidence 💌	Recall 💌	F-measure 💌	Lift 🖵	Convic 🔽 n	
2	291	2pct_Milk	& Onions &	Ramen_Noodles			Wheat_Brea	d	4	0.0176	0.8889	0.2286	0.3636	11.5217	6.6862	- 1
3	290	Sugar_Coo	okies & Sweet	_Relish & White	_Bread		Toothpaste		4	0.0154	0.913	0.1944	0.3206	11.506	7.6669	_
4	286	Graham_C	rackers & Po	tato_Chips			Toothpaste		3	0.0154	0.875	0.1944	0.3182	11.0266	5.9802	- 1
5	287	2pct_Milk	& Orange_Ju	ice & Potato_Ch	ips		Toothpaste		4	0.0154	0.875	0.1944	0.3182	11.0266	5.9802	
6	288	2pct_Milk	& Ravioli &	Sweet_Relish			Toothpaste		4	0.0154	0.875	0.1944	0.3182	11.0266	5.9802	_
7	289	Plums & S	weet_Relish	& White_Bread			Toothpaste		4	0.0154	0.875	0.1944	0.3182	11.0266	5.9802	_
8	285	Apples & I	Hot_Dogs				Toothpaste		3	0.0176	0.8571	0.2222	0.3529	10.8016	5.5202	- 1
9	283	Hot_Dog_E	Buns & Hot_D	ogs & Potatoes			Sweet_Relish	1	4	0.0154	0.913	0.181	0.3022	10.7125	7.618	- 1
10	284	Hot_Dogs	& Potatoes 8	Toothpaste			Sweet_Relish	1	4	0.0169	0.8846	0.1983	0.3239	10.379	6.3991	- 1
11	280	Bologna &	Eggs & Swee	t_Relish			Potatoes		4	0.0169	0.8846	0.1949	0.3194	10.2031	6.3888	
12	281	Chicken_S	oup & Eggs				Sweet_Relish	1	3	0.0176	0.8571	0.2069	0.3333	10.0567	5.485	_
13	282	2pct_Milk	& Cream_Ch	eese & Potatoes			Sweet_Relish	1	4	0.0169	0.8519	0.1983	0.3217	9.9946	5.3021	
	P PI	dataset	reull	2												
Pret													LUI EU 90 %	9	() ()	• .::

The rule with the highest LIFT is:

IF (2pct\_Milk & Onions & Ramen\_Noodles) THEN (Wheat\_Bread) [LIFT = 11.5217]

#### 4.3 Filtering according to a numerical condition

We want to display only the rules with a conviction > 6.

es données 👻	<ul> <li>↓</li> <li>↓</li></ul>	Grouj Disso	per ▼ 🃲	?₄ Solveur			
enarios *		sous-	total				
		Plar	n @	Analyse			
F		G	Н	- I			
Confidence 💌	Rec	all 💽	🖌 F-measu	re 💌 Lift 🛛	🕂 Convic 💌		
0.8889	) (	A↓	Trier du pl	u <u>s</u> petit au plu	is grand		
0.913	3 (	Z↓	Trier du pl	us grand au p	lus petit		
0.875	5 (	A 1	Trier par c	ouleur		Est égal à	Filtre automatique personnalisé
0.875	5 (	<b>.</b>	Effective last	cities de la Cerre		Ert différent de	Afficher les lignes dans lesguelles :
0.875	5 (	4%	Etta <u>c</u> er le i	riitre de « Con	viction »	Est differe <u>n</u> t de	Conviction
	_		F <u>i</u> ltrer par	couleur	Þ	<u>S</u> upérieur à	est supérieur ou égal à 🔻 6 🔍
			<u>F</u> iltres nur	mériques	· · · ·	Supérieur <u>o</u> u égal à	● <u>Et</u> ○ <u>Q</u> u
						Inférieur à	<b>v</b>
						Inférieur ou égal à	
						Entre	Utilisez ? pour représenter un caractère Utilisez * pour représenter une série de caractères
						10 premiers	OK Annuler
						Au-dessus de la moyenne	
						En dess <u>o</u> us de la moyenne	
						<u>F</u> iltre personnalisé	

To do this, we click on the corresponding arrow and we select the option "Number Filters / Greater than or Equal to...". We specify the threshold value 6.

#### Tanagra Tutorial

<b>.</b>	₩ → ) ≠			market_b	asket.	dsx - Micro	osoft Excel		_				x
Accueil	Insertion	Mise en page	Formules	Données	Rév	ision /	Affichage	Développeur	Comple	éments		<ul> <li> <ul> <li></li></ul></li></ul>	∍ x
Données externes v	Connexi Propriét Modifie	ions tés r les liens d'accès ons	$ \begin{array}{c c} \underline{A} & \underline{A} & \underline{Z} \\ \underline{Z} & \underline{A} \\ \underline{Z} & \underline{A} \\ \underline{A} & \underline{Trier} \end{array} $	Filtrer	facer éappliq vancé	uer Con	vertir Suppr les dou	rimer Jolons 💀 🔻	Grouper     Grouper     Dissocie     Sous-to     Plan	r + 0∰ <b>?₀</b> sr + "∰ tal	Solveur		
B3	<del>-</del> (9	<i>f</i> <sub>≭</sub> Suga	ar_Cookies &	Sweet_Relis	h & V	/hite_Bre	ad						×
A		В		С		D	E	F	G	Н	1	J	
1 Id 💌 Antecede	nt			Consequent	-	Length 💌	Support 💌	Confidence	🖌 Recall 🔽	F-measure	🔻 Lift 🛛 🤄	🕂 Convic 🖓	h 📄
2 291 2pct_Mil	k & Onions &	Ramen_Noodles		Wheat_Bread	b	4	0.0176	0.888	9 0.2286	0.363	36 11.521	.7 6.6862	
3 290 Sugar_Co	okies & Swee	et_Relish & White	_Bread	Toothpaste		4	0.0154	0.91	3 0.1944	0.320	06 11.50	6 7.6669	
9 283 Hot_Dog	Buns & Hot_	Dogs & Potatoes		Sweet_Relish	1	4	0.0154	0.91	3 0.181	0.302	10.712	5 7.618	
10 284 Hot_Dogs	s & Potatoes &	& Toothpaste		Sweet_Relish		4	0.0169	0.884	6 0.1983	0.323	39 10.37	9 6.3991	
11 280 Bologna	& Eggs & Swe	et_Relish		Potatoes		4	0.0169	0.884	6 0.1949	0.319	94 10.203	6.3888	-
H + + H datase	t Fouil1	*1					I 4	1				( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( ) ( )	
P 267 enregistr	ement(s) trouv	ré(s) sur 510								90 %	Θ	0	€ .:i

267 out of 510 rules meet this condition.

#### 4.4 Filtering based on the consequent of the rule

**Among these rules,** we search the ones with the consequent "Purchase of (Potato\_Chips)". We click the arrow in the header of the column and we select the corresponding item.

	c 🥾
	💌 Consequent 🦷
₽↓	Trier de <u>A</u> à Z
Ă↑	Trier de <u>Z</u> à A
	Trier par couleur
K	Effa <u>c</u> er le filtre de « Consequent »
	F <u>i</u> ltrer par couleur ▶
	<u>F</u> iltres textuels
	Sélectionner tout)     2pct_Milk     Eggs     Hot_Dogs     Potatoes     Sweet_Relish     Toothpaste     Wheat_Bread     White_Bread
	OK Annuler

6 rules are highlighted.

		B3 ▼ ( ≸ Sugar_Cookies & Swe	et_Relish & Wl	nite	e_Bread						3	
	А	В	С		D	E	F	G	н	1	J	
1	Id 💌	Antecedent 🗸	Consequent	<b>7</b>	Length 💌	Support 💌	Confidence 💌	Recall 💌	F-measure 💌	Lift 🖵	Convic 🖓 n	
22	268	2pct_Milk & 98pct_Fat_Free_Hamburger & Toothpaste	Potato_Chips		4	0.0184	0.8929	0.188	0.3106	9.1367	6.7627	
25	276	98pct_Fat_Free_Hamburger & Garlic & White_Bread	Potato_Chips		4	0.0176	0.8889	0.1805	0.3	9.0961	6.5372	
26	263	Apples & Canned_Tuna	Potato_Chips		3	0.0162	0.88	0.1654	0.2785	9.0051	6.0864	
27	265	2pct_Milk & 98pct_Fat_Free_Hamburger & Onions	Potato_Chips		4	0.0162	0.88	0.1654	0.2785	9.0051	6.0864	
28	270	2pct_Milk & Apples & Potatoes	Potato_Chips		4	0.0162	0.88	0.1654	0.2785	9.0051	6.0864	
29	278	98pct_Fat_Free_Hamburger & Toothpaste & Wheat_Bread	Potato_Chips		4	0.0162	0.88	0.1654	0.2785	9.0051	6.0864	
14	▲ ► ► dataset Feuil1 💱											
	6 e	nregistrement(s) trouvé(s) sur 510						<b>= =</b>	90 % 😑	Ū.	+	

## 4.1 Filtering based on the antecedent of the rule

**Among these rules**, we want to highlight the rules for which the antecedent contains the "2pct\_Milk" item. We use the **Text Filters** tool, we set the condition.



3 rules are now displayed.

		B3 ▼ ( \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$\$ \$	et_Relish & Whit	e_Bread						×
	Α	В	С	D	E	F	G	н	1	J
1	Id 💌	Antecedent 🖓	Consequent 🛛 🖓	🛛 Length 💌	Support 💌	Confidence 💌	Recall 💌	F-measure 💌	Lift 斗	Convic 🖓 🗧
22	268	2pct_Milk & 98pct_Fat_Free_Hamburger & Toothpaste	Potato_Chips	4	0.0184	0.8929	0.188	0.3106	9.1367	6.7627
27	265	2pct_Milk & 98pct_Fat_Free_Hamburger & Onions	Potato_Chips	4	0.0162	0.88	0.1654	0.2785	9.0051	6.0864
28	270	2pct_Milk & Apples & Potatoes	Potato_Chips	4	0.0162	0.88	0.1654	0.2785	9.0051	6.0864
II I	► • H	dataset Fouil1 /			[ ∢ _		1111			▶ [
	3 er	nregistrement(s) trouvé(s) sur 510 🚬						90 % 🕒	)	

We can thus multiply the combinations to highlight the rules that best meet the specifications of our study. To be honest, this system is only really operational if we deal with a moderate-size rules base. But Excel is easy to use. It is one of the data miner favorite tools for a long time (KDnuggets Polls, <u>Top Analytics / Data Science Tools</u>, May 2017). We note in this tutorial that its functionalities are attractive for the processing of rule bases.

# 5 Conclusion

ARS is an academic software for association rule mining. It is included into the SIPINA distribution. In this tutorial, we describe its combined use with Excel. Among other things, the opportunities about the post-processing of the mined rules is particularly attractive.