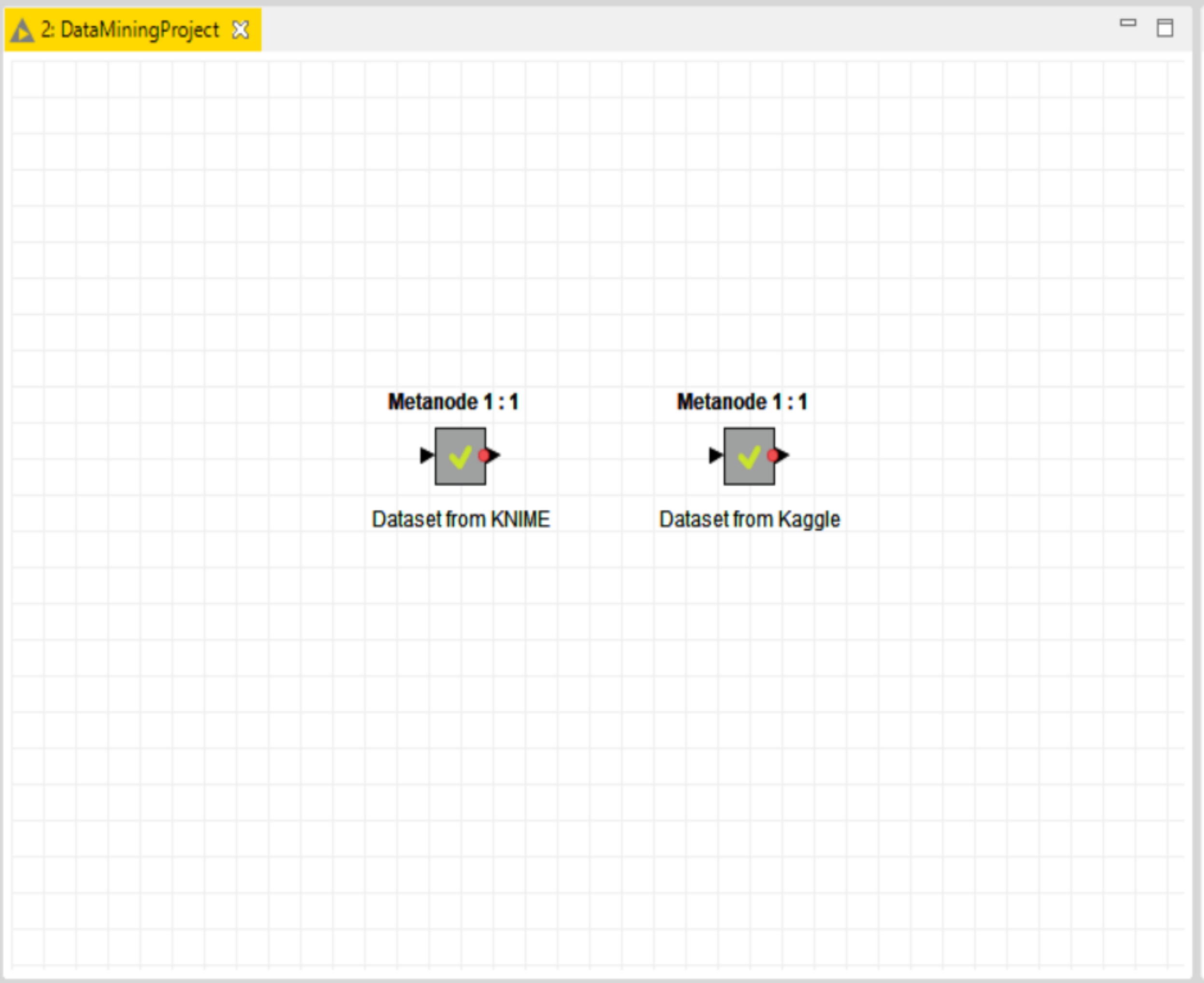
ABSTRACT

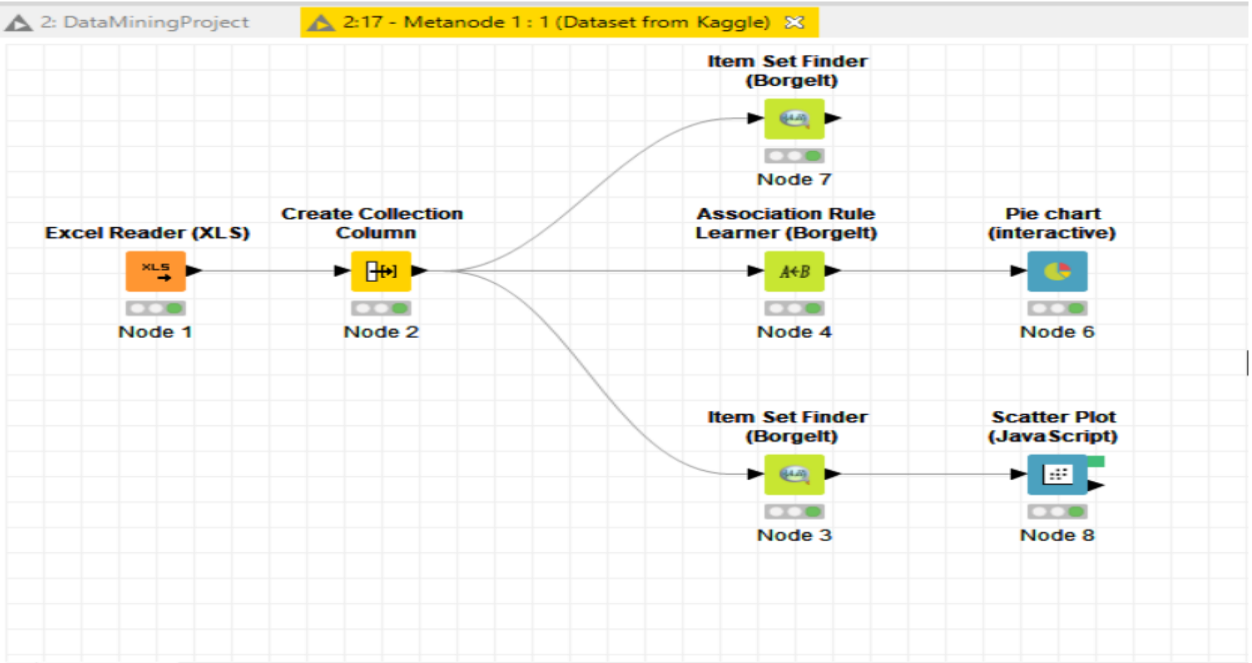
Main purpose of working on this project is analyzing data sets which contain market data to apply some algorithms such as Apriori, FPGrowth, TANIMOTO and more to find frequent item sets to show results through pie chart and scatter plot on KNIME.

STEPS OF THE PROJECT

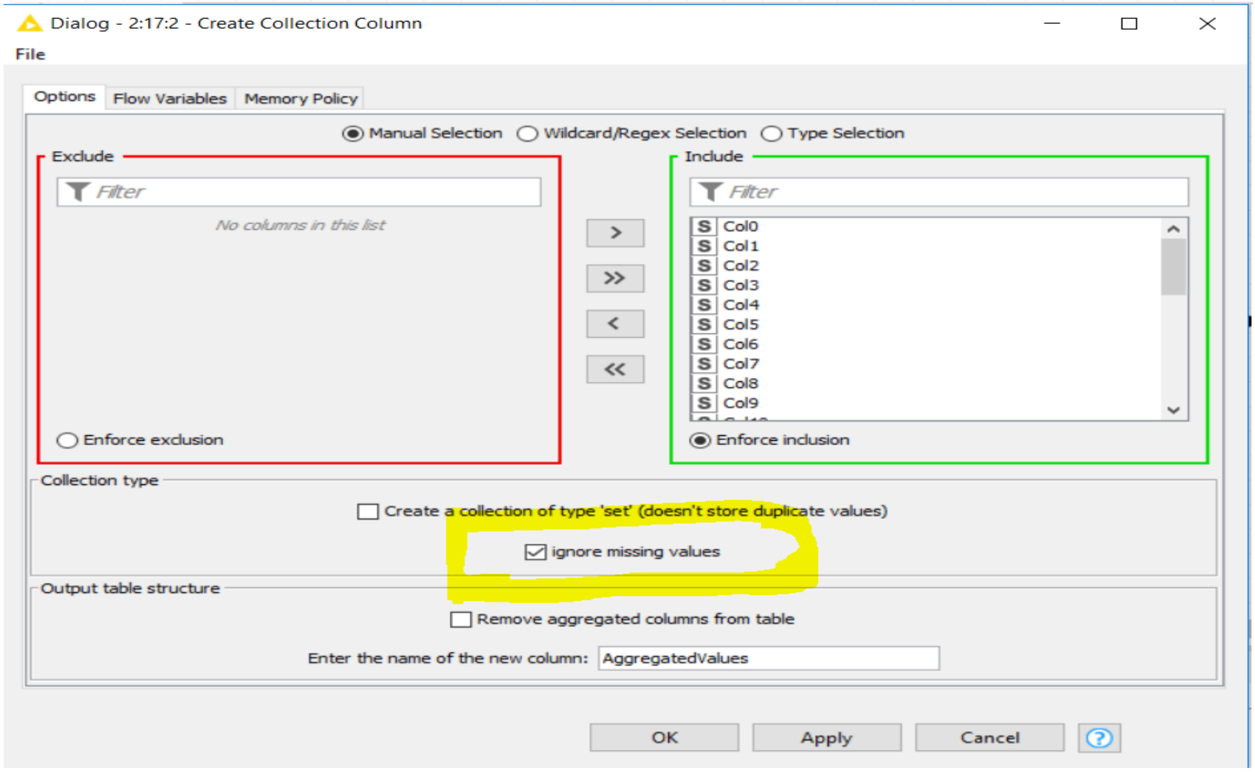
We used two different data sets to get better results. The first data set that we applied algorithms on from Kaggle (https://www.kaggle.com/apmonisha08/market-basket-analysis) and the second data set from KNIME's data set example for Basket Analysis. We created two different metanodes not to be distracted while working on data sets from different sources.



Firstly, we are going to explain our project which we used the data set from Kaggle. In the figure below, you will be able to see the steps of the project such as reading the data as an Excel file, then we created data collection (like an array) to apply algorithms on our data set through “Create Collection Column” and as a step of preprocessing and due to have a lot of missing values, we ignored missing values not to get wrong/worse results from our data analysis.



2: DataMiningProject
2:17 - Metanode 1 : 1 (Dataset from Kaggle)
Excel Reader (XLS)
Node 1
Create Collection Column
Node 2
Item Set Finder (Borgelt)
Node 7
Association Rule Learner (Borgelt)
Node 4
Pie chart (interactive)
Node 6
Item Set Finder (Borgelt)
Node 3
Scatter Plot (JavaScript)
Node 8



● Manual Selection ○ Wildcard/Regex Selection ○ Type Selection

Exclude

Filter

No columns in this list

○ Enforce exclusion

Include

Filter

S Col0
S Col1
S Col2
S Col3
S Col4
S Col5
S Col6
S Col7
S Col8
S Col9

● Enforce inclusion

Collection type

□ Create a collection of type 'set' (doesn't store duplicate values)

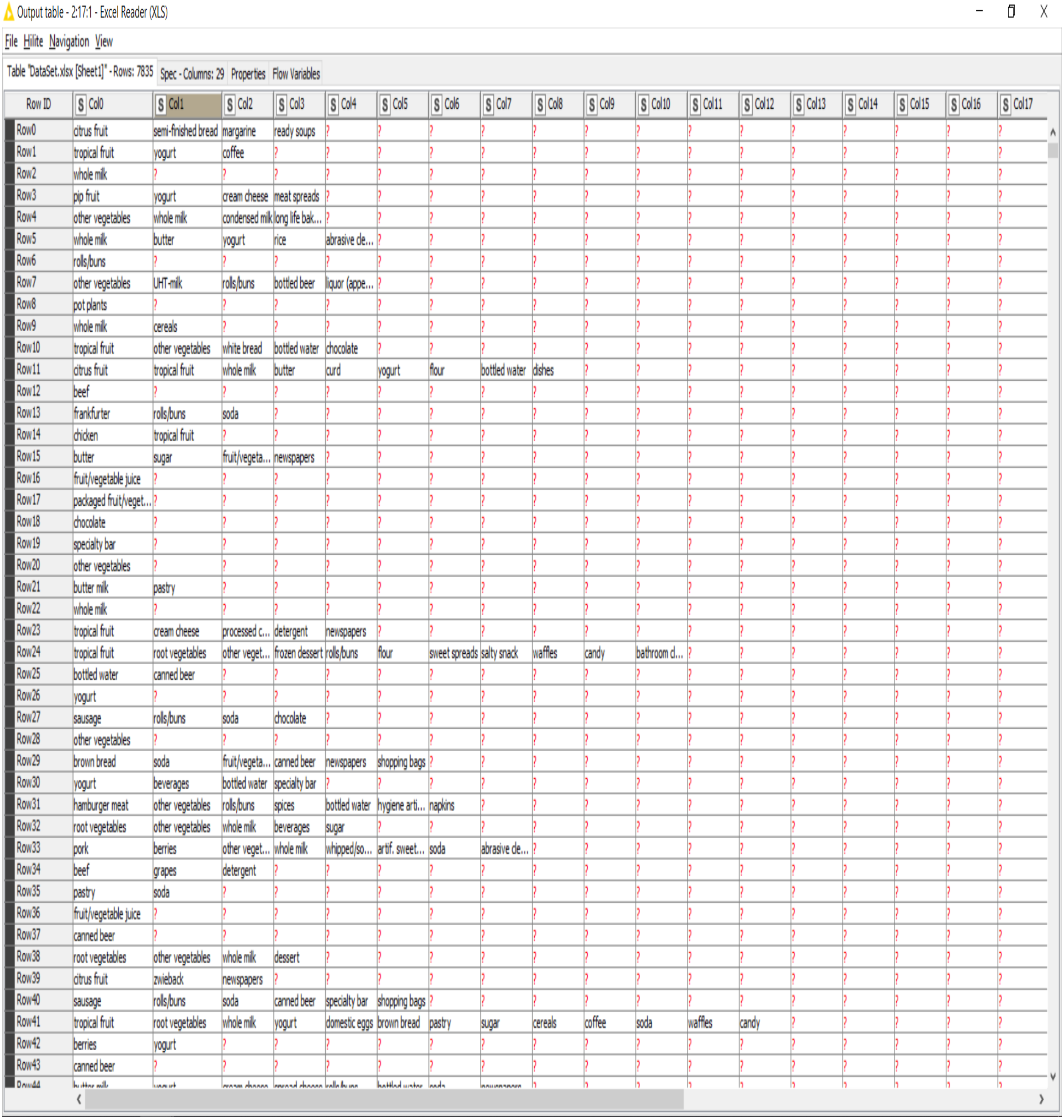
☑ ignore missing values

Output table structure

□ Remove aggregated columns from table

Enter the name of the new column: AggregatedValues

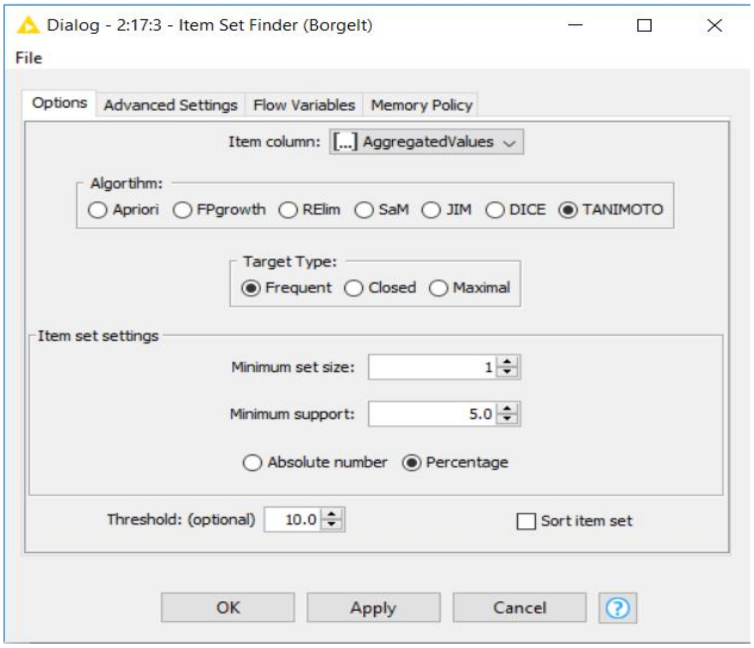
Before ignoring the missing values, our data set looked like in the figure below and it might cause have wrong results if we also add them our analysis.



Output table - 2:17:1 - Excel Reader (XLS)
File Hilite Navigation View
Table "DataSet.xlsx [Sheet1]" - Rows: 7835
Spec - Columns: 29
Properties
Flow Variables

Row ID	S Col0	S Col1	S Col2	S Col3	S Col4	S Col5	S Col6	S Col7	S Col8	S Col9	S Col10	S Col11	S Col12	S Col13	S Col14	S Col15	S Col16	S Col17
Row0	citrus fruit	semi-finished bread	margarine	ready soups	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row1	tropical fruit	yogurt	coffee	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row2	whole milk	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row3	pip fruit	yogurt	cream cheese	meat spreads	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row4	other vegetables	whole milk	condensed milk	long life bak...	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row5	whole milk	butter	yogurt	rice	abrasive cle...	?	?	?	?	?	?	?	?	?	?	?	?	?
Row6	rolls/buns	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row7	other vegetables	UHT-milk	rolls/buns	bottled beer	liquor (appe...	?	?	?	?	?	?	?	?	?	?	?	?	?
Row8	pot plants	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row9	whole milk	cereals	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row10	tropical fruit	other vegetables	white bread	bottled water	chocolate	?	?	?	?	?	?	?	?	?	?	?	?	?
Row11	citrus fruit	tropical fruit	whole milk	butter	curd	yogurt	flour	bottled water	dishes	?	?	?	?	?	?	?	?	?
Row12	beef	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row13	frankfurter	rolls/buns	soda	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row14	chicken	tropical fruit	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row15	butter	sugar	fruit/vegeta...	newspapers	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row16	fruit/vegetable juice	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row17	packaged fruit/veget...	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row18	chocolate	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row19	specialty bar	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row20	other vegetables	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row21	butter milk	pastry	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row22	whole milk	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row23	tropical fruit	cream cheese	processed c...	detergent	newspapers	?	?	?	?	?	?	?	?	?	?	?	?	?
Row24	tropical fruit	root vegetables	other veget...	frozen dessert	rolls/buns	flour	sweet spreads	salty snack	waffles	candy	bathroom cl...	?	?	?	?	?	?	?
Row25	bottled water	canned beer	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row26	yogurt	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row27	sausage	rolls/buns	soda	chocolate	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row28	other vegetables	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row29	brown bread	soda	fruit/vegeta...	canned beer	newspapers	shopping bags	?	?	?	?	?	?	?	?	?	?	?	?
Row30	yogurt	beverages	bottled water	specialty bar	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row31	hamburger meat	other vegetables	rolls/buns	spices	bottled water	hygiene arti...	napkins	?	?	?	?	?	?	?	?	?	?	?
Row32	root vegetables	other vegetables	whole milk	beverages	sugar	?	?	?	?	?	?	?	?	?	?	?	?	?
Row33	pork	berries	other veget...	whole milk	whipped/so...	artif. sweet...	soda	abrasive cle...	?	?	?	?	?	?	?	?	?	?
Row34	beef	grapes	detergent	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row35	pastry	soda	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row36	fruit/vegetable juice	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row37	canned beer	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row38	root vegetables	other vegetables	whole milk	dessert	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row39	citrus fruit	zwieback	newspapers	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row40	sausage	rolls/buns	soda	canned beer	specialty bar	shopping bags	?	?	?	?	?	?	?	?	?	?	?	?
Row41	tropical fruit	root vegetables	whole milk	yogurt	domestic eggs	brown bread	pastry	sugar	cereals	coffee	soda	waffles	candy	?	?	?	?	?
Row42	berries	yogurt	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row43	canned beer	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Row44	butter milk	yogurt	cream cheese	spread cheese	rolls/buns	bottled water	soda	newspapers	?	?	?	?	?	?	?	?	?	?

After we applied data preprocesses, then we applied FPGrowth and TANIMOTO algorithms to see the results of analysis. We have chosen minimum support values as 5% and 4% and minimum set size as 1.


Dialog - 2:17:3 - Item Set Finder (Borgelt)

File

Options | Advanced Settings | Flow Variables | Memory Policy

Item column: [...] AggregatedValues

Algorthim: ○ Apriori ○ FPgrowth ○ RElim ○ SaM ○ JIM ○ DICE ● TANIMOTO

Target Type: ● Frequent ○ Closed ○ Maximal

Item set settings

Minimum set size: 1

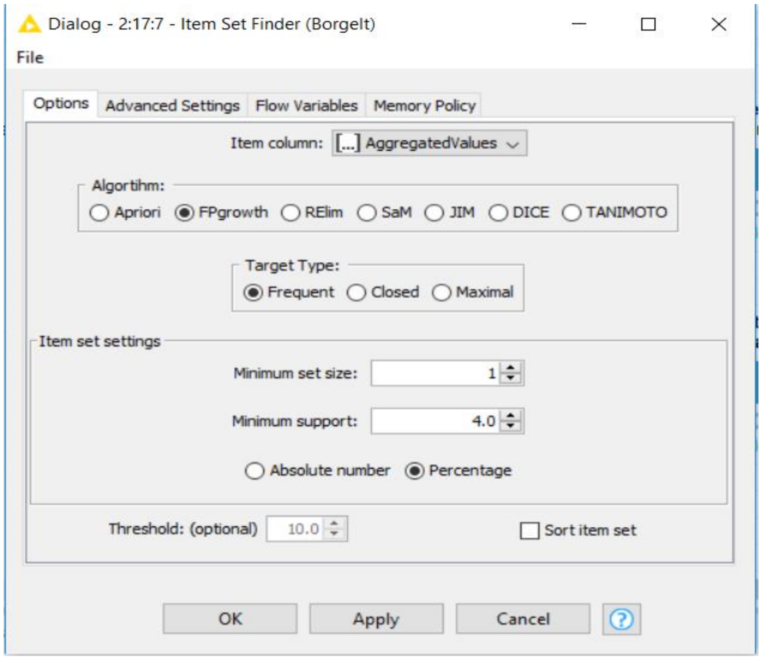
Minimum support: 5.0

○ Absolute number ● Percentage

Threshold: (optional) 10.0

☐ Sort item set

OK | Apply | Cancel | ?


Dialog - 2:17:7 - Item Set Finder (Borgelt)

File

Options | Advanced Settings | Flow Variables | Memory Policy

Item column: [...] AggregatedValues

Algorthim: ○ Apriori ● FPgrowth ○ RElim ○ SaM ○ JIM ○ DICE ○ TANIMOTO

Target Type: ● Frequent ○ Closed ○ Maximal

Item set settings

Minimum set size: 1

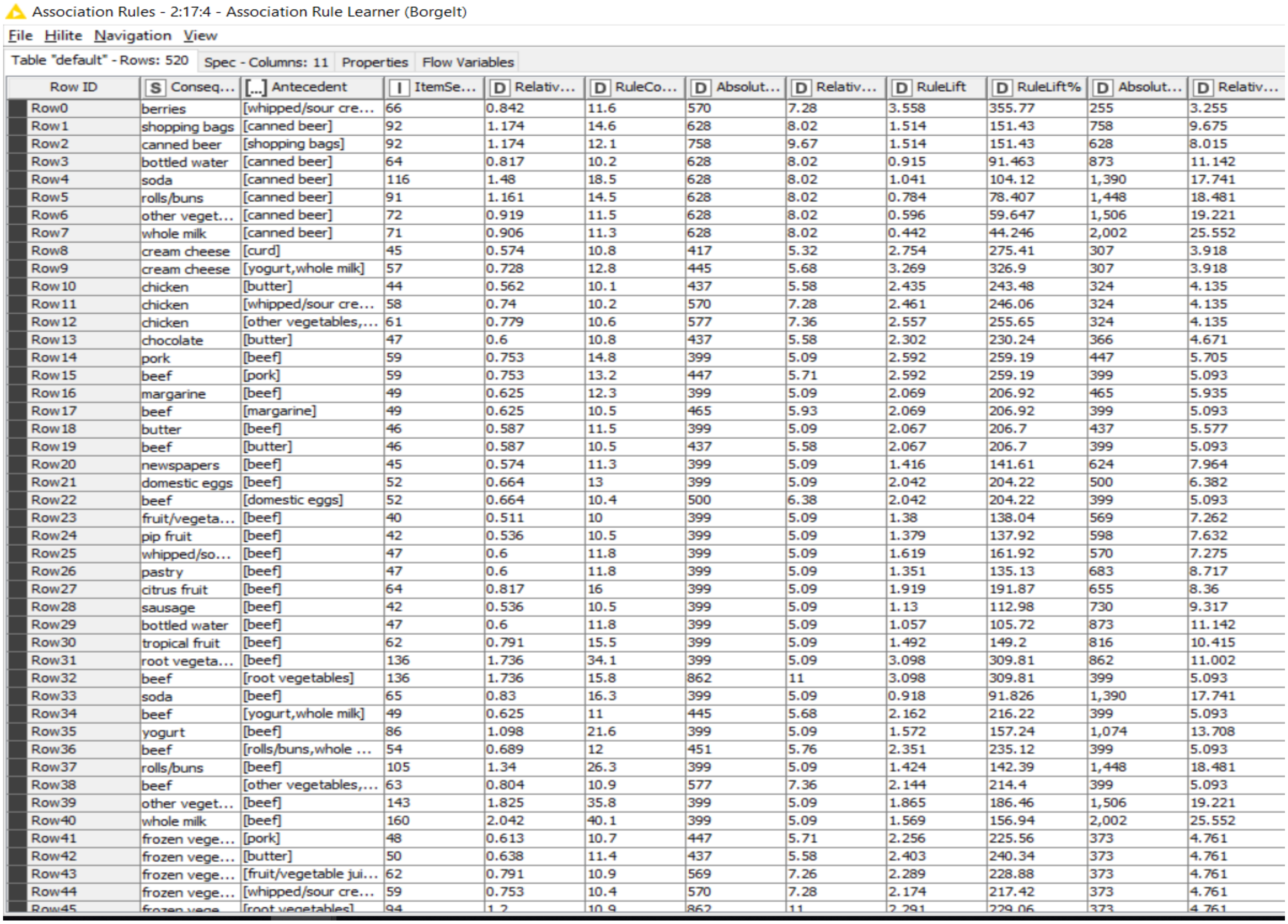
Minimum support: 4.0

○ Absolute number ● Percentage

Threshold: (optional) 10.0

☐ Sort item set

OK | Apply | Cancel | ?

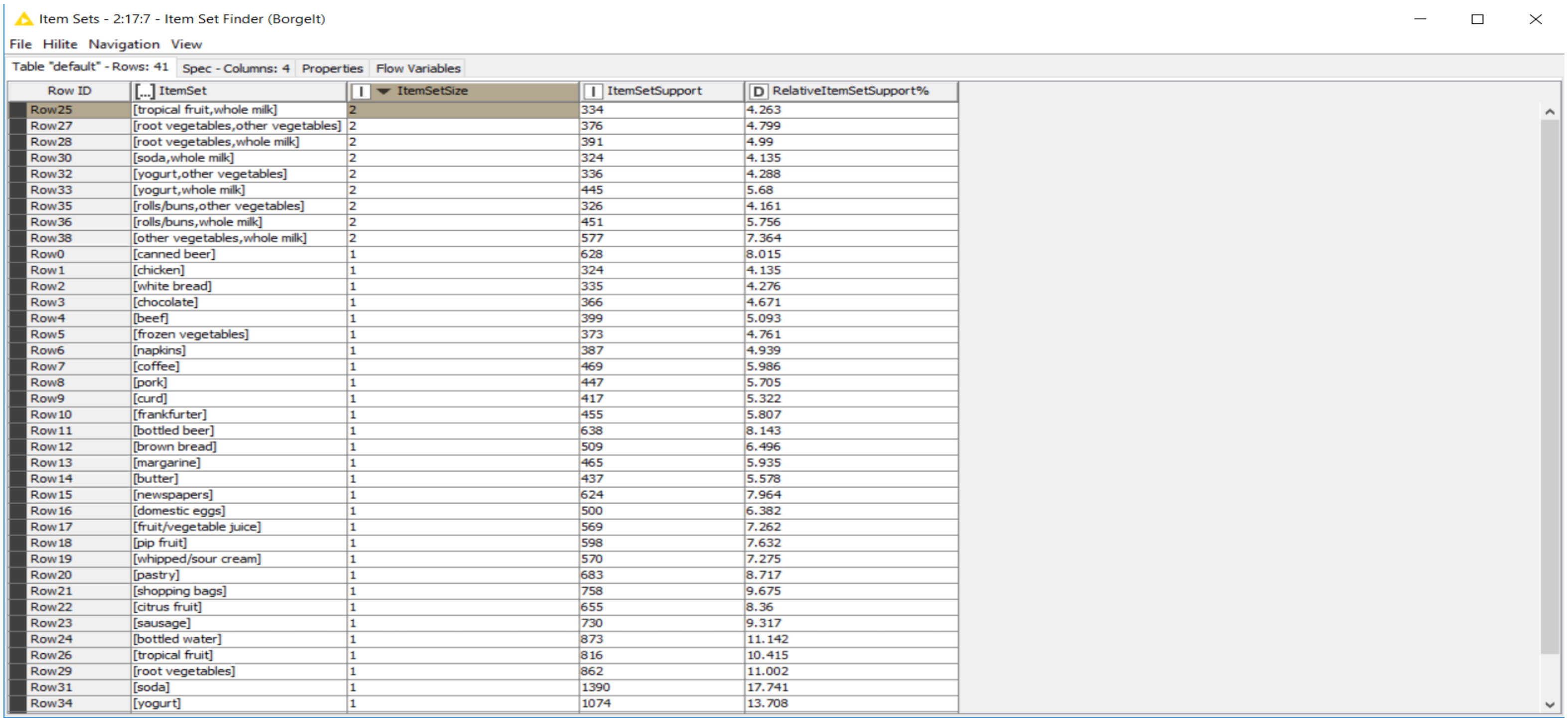

Association Rules - 2:17:4 - Association Rule Learner (Borgelt)

File Hilite Navigation View

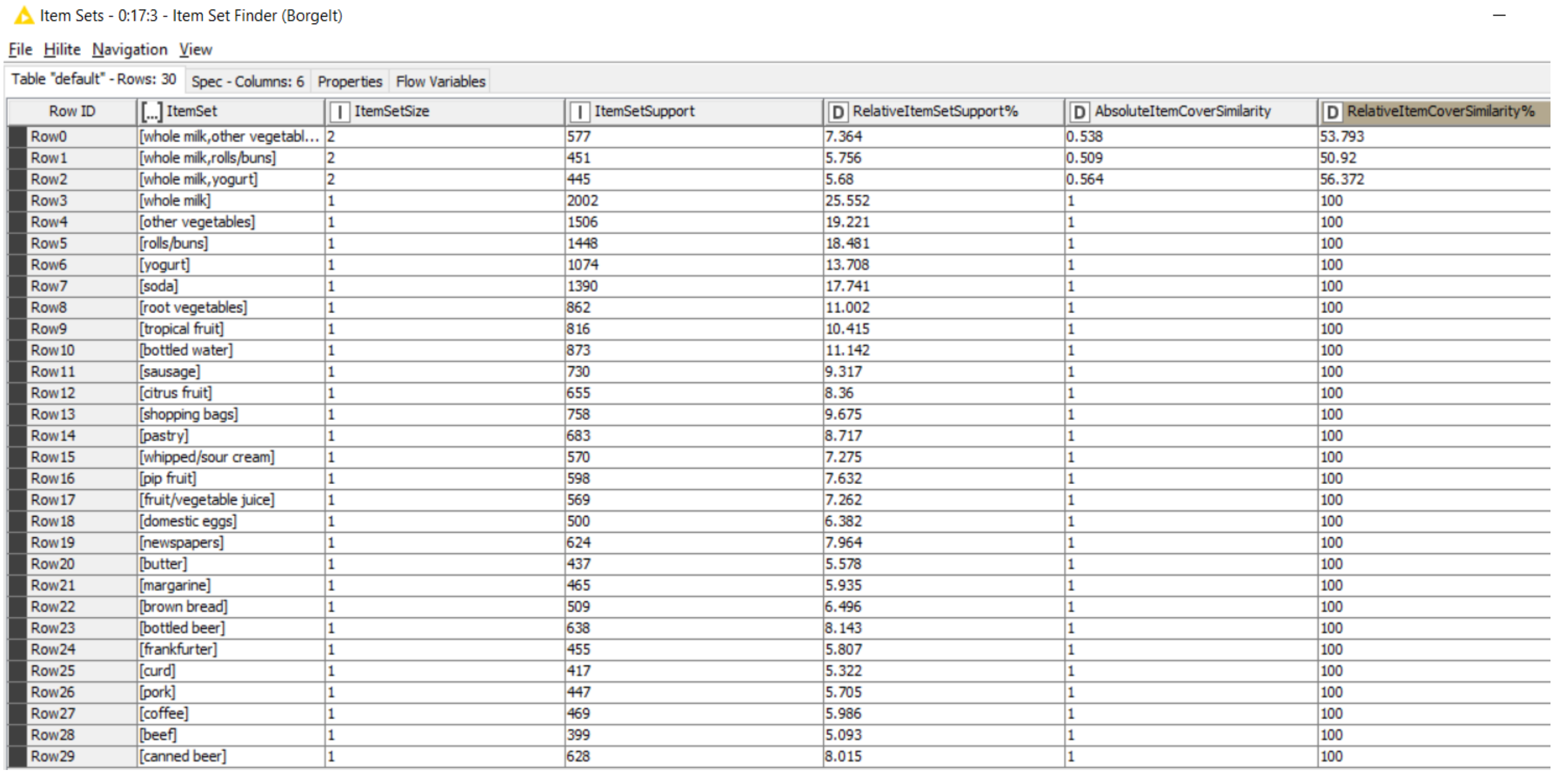
Table "default" - Rows: 520 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	S Conseq...	[...] Antecedent	I ItemSe...	D Relativ...	D RuleCo...	D Absolut...	D Relativ...	D RuleLift	D RuleLift%	D Absolut...	D Relativ...
Row0	berries	[whipped/sour cre...	66	0.842	11.6	570	7.28	3.558	355.77	255	3.255
Row1	shopping bags	[canned beer]	92	1.174	14.6	628	8.02	1.514	151.43	758	9.675
Row2	canned beer	[shopping bags]	92	1.174	12.1	758	9.67	1.514	151.43	628	8.015
Row3	bottled water	[canned beer]	64	0.817	10.2	628	8.02	0.915	91.463	873	11.142
Row4	soda	[canned beer]	116	1.48	18.5	628	8.02	1.041	104.12	1,390	17.741
Row5	rolls/buns	[canned beer]	91	1.161	14.5	628	8.02	0.784	78.407	1,448	18.481
Row6	other veget...	[canned beer]	72	0.919	11.5	628	8.02	0.596	59.647	1,506	19.221
Row7	whole milk	[canned beer]	71	0.906	11.3	628	8.02	0.442	44.246	2,002	25.552
Row8	cream cheese	[curd]	45	0.574	10.8	417	5.32	2.754	275.41	307	3.918
Row9	cream cheese	[yogurt,whole milk]	57	0.728	12.8	445	5.68	3.269	326.9	307	3.918
Row10	chicken	[butter]	44	0.562	10.1	437	5.58	2.435	243.48	324	4.135
Row11	chicken	[whipped/sour cre...	58	0.74	10.2	570	7.28	2.461	246.06	324	4.135
Row12	chicken	[other vegetables,...	61	0.779	10.6	577	7.36	2.557	255.65	324	4.135
Row13	chocolate	[butter]	47	0.6	10.8	437	5.58	2.302	230.24	366	4.671
Row14	pork	[beef]	59	0.753	14.8	399	5.09	2.592	259.19	447	5.705
Row15	beef	[pork]	59	0.753	13.2	447	5.71	2.592	259.19	399	5.093
Row16	margarine	[beef]	49	0.625	12.3	399	5.09	2.069	206.92	465	5.935
Row17	beef	[margarine]	49	0.625	10.5	465	5.93	2.069	206.92	399	5.093
Row18	butter	[beef]	46	0.587	11.5	399	5.09	2.067	206.7	437	5.577
Row19	beef	[butter]	46	0.587	10.5	437	5.58	2.067	206.7	399	5.093
Row20	newspapers	[beef]	45	0.574	11.3	399	5.09	1.416	141.61	624	7.964
Row21	domestic eggs	[beef]	52	0.664	13	399	5.09	2.042	204.22	500	6.382
Row22	beef	[domestic eggs]	52	0.664	10.4	500	6.38	2.042	204.22	399	5.093
Row23	fruit/vegeta...	[beef]	40	0.511	10	399	5.09	1.38	138.04	569	7.262
Row24	pip fruit	[beef]	42	0.536	10.5	399	5.09	1.379	137.92	598	7.632
Row25	whipped/so...	[beef]	47	0.6	11.8	399	5.09	1.619	161.92	570	7.275
Row26	pastry	[beef]	47	0.6	11.8	399	5.09	1.351	135.13	683	8.717
Row27	citrus fruit	[beef]	64	0.817	16	399	5.09	1.919	191.87	655	8.36
Row28	sausage	[beef]	42	0.536	10.5	399	5.09	1.13	112.98	730	9.317
Row29	bottled water	[beef]	47	0.6	11.8	399	5.09	1.057	105.72	873	11.142
Row30	tropical fruit	[beef]	62	0.791	15.5	399	5.09	1.492	149.2	816	10.415
Row31	root vegeta...	[beef]	136	1.736	34.1	399	5.09	3.098	309.81	862	11.002
Row32	beef	[root vegetables]	136	1.736	15.8	862	11	3.098	309.81	399	5.093
Row33	soda	[beef]	65	0.83	16.3	399	5.09	0.918	91.826	1,390	17.741
Row34	beef	[yogurt,whole milk]	49	0.625	11	445	5.68	2.162	216.22	399	5.093
Row35	yogurt	[beef]	86	1.098	21.6	399	5.09	1.572	157.24	1,074	13.708
Row36	beef	[rolls/buns,whole ...	54	0.689	12	451	5.76	2.351	235.12	399	5.093
Row37	rolls/buns	[beef]	105	1.34	26.3	399	5.09	1.424	142.39	1,448	18.481
Row38	beef	[other vegetables,...	63	0.804	10.9	577	7.36	2.144	214.4	399	5.093
Row39	other veget...	[beef]	143	1.825	35.8	399	5.09	1.865	186.46	1,506	19.221
Row40	whole milk	[beef]	160	2.042	40.1	399	5.09	1.569	156.94	2,002	25.552
Row41	frozen vege...	[pork]	48	0.613	10.7	447	5.71	2.256	225.56	373	4.761
Row42	frozen vege...	[butter]	50	0.638	11.4	437	5.58	2.403	240.34	373	4.761
Row43	frozen vege...	[fruit/vegetable jui...	62	0.791	10.9	569	7.26	2.289	228.88	373	4.761
Row44	frozen vege...	[whipped/sour cre...	59	0.753	10.4	570	7.28	2.174	217.42	373	4.761
Row45	frozen vege...	[root vegetables]	94	1.2	10.9	862	11	2.291	229.06	373	4.761

In the results below, we are able to get some of information according to the result. Firstly, let’s consider “ItemSetSize” and sort them descending order. If we look Row 25,Row 27,Row 28 and Row 30,we are able to say that if the market sells tropical fruit and whole milk together,4.263 percent of customers will buy and if we check “ItemSetSupport” for Row 25 is exist 334 times in all data set as a frequent item set. We can consider what we need to see on this results and show them in the results section. First photo represents the results of FPGrowth and second is TANIMOTO algorithm shows “similarity” which is used to check the model’s robustness.

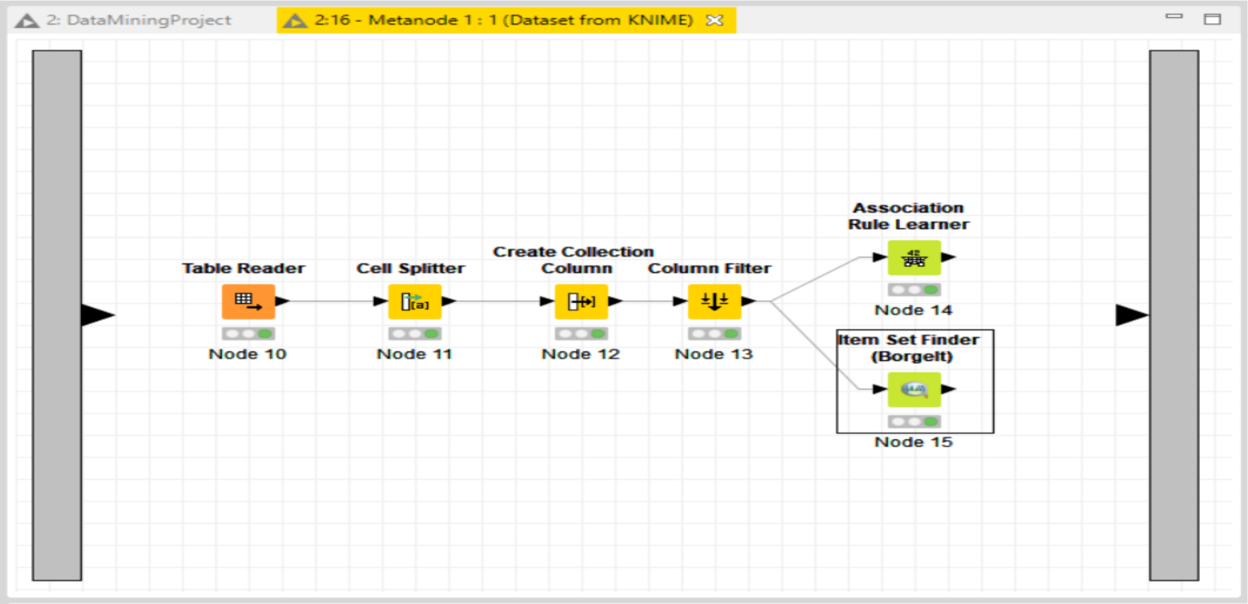

Item Sets - 2:17:7 - Item Set Finder (Borgelt)
File Hilite Navigation View
Table "default" - Rows: 41
Spec - Columns: 4
Properties
Flow Variables

Row ID	[...] ItemSet	I ItemSetSize	I ItemSetSupport	D RelativeItemSetSupport%
Row25	[tropical fruit,whole milk]	2	334	4.263
Row27	[root vegetables,other vegetables]	2	376	4.799
Row28	[root vegetables,whole milk]	2	391	4.99
Row30	[soda,whole milk]	2	324	4.135
Row32	[yogurt,other vegetables]	2	336	4.288
Row33	[yogurt,whole milk]	2	445	5.68
Row35	[rolls/buns,other vegetables]	2	326	4.161
Row36	[rolls/buns,whole milk]	2	451	5.756
Row38	[other vegetables,whole milk]	2	577	7.364
Row0	[canned beer]	1	628	8.015
Row1	[chicken]	1	324	4.135
Row2	[white bread]	1	335	4.276
Row3	[chocolate]	1	366	4.671
Row4	[beef]	1	399	5.093
Row5	[frozen vegetables]	1	373	4.761
Row6	[napkins]	1	387	4.939
Row7	[coffee]	1	469	5.986
Row8	[pork]	1	447	5.705
Row9	[curd]	1	417	5.322
Row10	[frankfurter]	1	455	5.807
Row11	[bottled beer]	1	638	8.143
Row12	[brown bread]	1	509	6.496
Row13	[margarine]	1	465	5.935
Row14	[butter]	1	437	5.578
Row15	[newspapers]	1	624	7.964
Row16	[domestic eggs]	1	500	6.382
Row17	[fruit/vegetable juice]	1	569	7.262
Row18	[pip fruit]	1	598	7.632
Row19	[whipped/sour cream]	1	570	7.275
Row20	[pastry]	1	683	8.717
Row21	[shopping bags]	1	758	9.675
Row22	[citrus fruit]	1	655	8.36
Row23	[sausage]	1	730	9.317
Row24	[bottled water]	1	873	11.142
Row26	[tropical fruit]	1	816	10.415
Row29	[root vegetables]	1	862	11.002
Row31	[soda]	1	1390	17.741
Row34	[yogurt]	1	1074	13.708

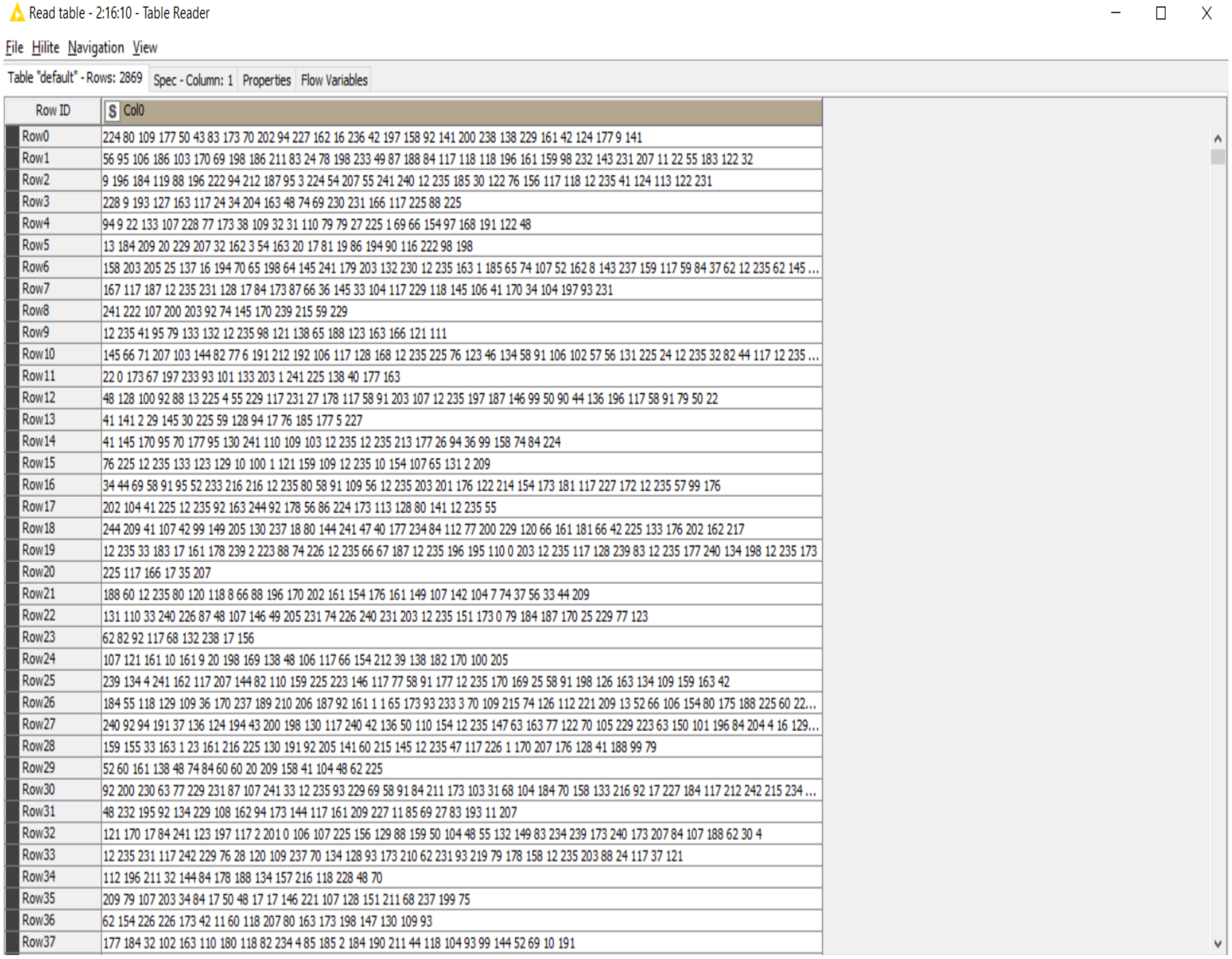

Item Sets - 0:17:3 - Item Set Finder (Borgelt)
File Hilite Navigation View
Table "default" - Rows: 30
Spec - Columns: 6
Properties
Flow Variables

Row ID	[...] ItemSet	I ItemSetSize	I ItemSetSupport	D RelativeItemSetSupport%	D AbsoluteItemCoverSimilarity	D RelativeItemCoverSimilarity%
Row0	[whole milk,other vegetabl...	2	577	7.364	0.538	53.793
Row1	[whole milk,rolls/buns]	2	451	5.756	0.509	50.92
Row2	[whole milk,yogurt]	2	445	5.68	0.564	56.372
Row3	[whole milk]	1	2002	25.552	1	100
Row4	[other vegetables]	1	1506	19.221	1	100
Row5	[rolls/buns]	1	1448	18.481	1	100
Row6	[yogurt]	1	1074	13.708	1	100
Row7	[soda]	1	1390	17.741	1	100
Row8	[root vegetables]	1	862	11.002	1	100
Row9	[tropical fruit]	1	816	10.415	1	100
Row10	[bottled water]	1	873	11.142	1	100
Row11	[sausage]	1	730	9.317	1	100
Row12	[citrus fruit]	1	655	8.36	1	100
Row13	[shopping bags]	1	758	9.675	1	100
Row14	[pastry]	1	683	8.717	1	100
Row15	[whipped/sour cream]	1	570	7.275	1	100
Row16	[pip fruit]	1	598	7.632	1	100
Row17	[fruit/vegetable juice]	1	569	7.262	1	100
Row18	[domestic eggs]	1	500	6.382	1	100
Row19	[newspapers]	1	624	7.964	1	100
Row20	[butter]	1	437	5.578	1	100
Row21	[margarine]	1	465	5.935	1	100
Row22	[brown bread]	1	509	6.496	1	100
Row23	[bottled beer]	1	638	8.143	1	100
Row24	[frankfurter]	1	455	5.807	1	100
Row25	[curd]	1	417	5.322	1	100
Row26	[pork]	1	447	5.705	1	100
Row27	[coffee]	1	469	5.986	1	100
Row28	[beef]	1	399	5.093	1	100
Row29	[canned beer]	1	628	8.015	1	100

Now, we are going to explain the second part of our project and it looks like in the figure below.

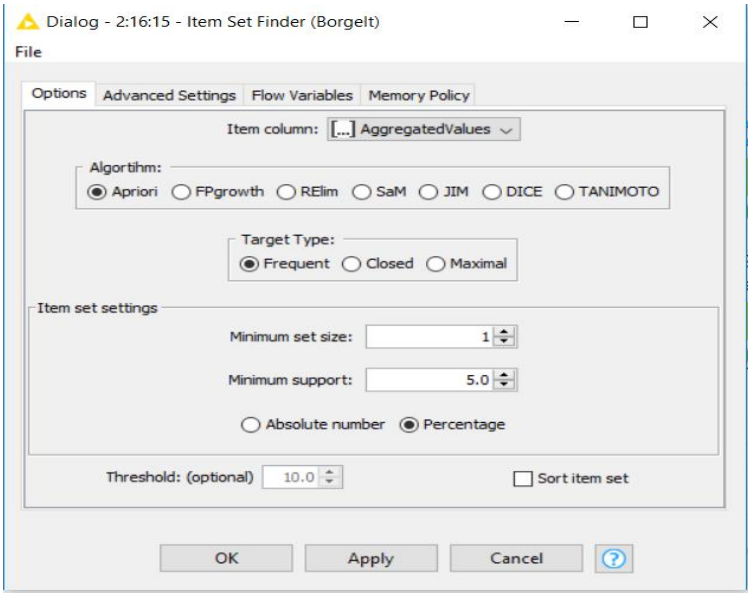


Our data set before applying preprocessing steps look like that :


Read table - 2:16:10 - Table Reader
File Hilite Navigation View
Table "default" - Rows: 2869
Spec - Column: 1
Properties
Flow Variables

Row ID	S Col0
Row0	224 80 109 177 50 43 83 173 70 202 94 227 162 16 236 42 197 158 92 141 200 238 138 229 161 42 124 177 9 141
Row1	56 95 106 186 103 170 69 198 186 211 83 24 78 198 233 49 87 188 84 117 118 118 196 161 159 98 232 143 231 207 11 22 55 183 122 32
Row2	9 196 184 119 88 196 222 94 212 187 95 3 224 54 207 55 241 240 12 235 185 30 122 76 156 117 118 12 235 41 124 113 122 231
Row3	228 9 193 127 163 117 24 34 204 163 48 74 69 230 231 166 117 225 88 225
Row4	94 9 22 133 107 228 77 173 38 109 32 31 110 79 79 27 225 1 69 66 154 97 168 191 122 48
Row5	13 184 209 20 229 207 32 162 3 54 163 20 17 81 19 86 194 90 116 222 98 198
Row6	158 203 205 25 137 16 194 70 65 198 64 145 241 179 203 132 230 12 235 163 1 185 65 74 107 52 162 8 143 237 159 117 59 84 37 62 12 235 62 145 ...
Row7	167 117 187 12 235 231 128 17 84 173 87 66 36 145 33 104 117 229 118 145 106 41 170 34 104 197 93 231
Row8	241 222 107 200 203 92 74 145 170 239 215 59 229
Row9	12 235 41 95 79 133 132 12 235 98 121 138 65 188 123 163 166 121 111
Row10	145 66 71 207 103 144 82 77 6 191 212 192 106 117 128 168 12 235 225 76 123 46 134 58 91 106 102 57 56 131 225 24 12 235 32 82 44 117 12 235 ...
Row11	22 0 173 67 197 233 93 101 133 203 1 241 225 138 40 177 163
Row12	48 128 100 92 88 13 225 4 55 229 117 231 27 178 117 58 91 203 107 12 235 197 187 146 99 50 90 44 136 196 117 58 91 79 50 22
Row13	41 141 2 29 145 30 225 59 128 94 17 76 185 177 5 227
Row14	41 145 170 95 70 177 95 130 241 110 109 103 12 235 12 235 213 177 26 94 36 99 158 74 84 224
Row15	76 225 12 235 133 123 129 10 100 1 121 159 109 12 235 10 154 107 65 131 2 209
Row16	34 44 69 58 91 95 52 233 216 216 12 235 80 58 91 109 56 12 235 203 201 176 122 214 154 173 181 117 227 172 12 235 57 99 176
Row17	202 104 41 225 12 235 92 163 244 92 178 56 86 224 173 113 128 80 141 12 235 55
Row18	244 209 41 107 42 99 149 205 130 237 18 80 144 241 47 40 177 234 84 112 77 200 229 120 66 161 181 66 42 225 133 176 202 162 217
Row19	12 235 33 183 17 161 178 239 2 223 88 74 226 12 235 66 67 187 12 235 196 195 110 0 203 12 235 117 128 239 83 12 235 177 240 134 198 12 235 173
Row20	225 117 166 17 35 207
Row21	188 60 12 235 80 120 118 8 66 88 196 170 202 161 154 176 161 149 107 142 104 7 74 37 56 33 44 209
Row22	131 110 33 240 226 87 48 107 146 49 205 231 74 226 240 231 203 12 235 151 173 0 79 184 187 170 25 229 77 123
Row23	62 82 92 117 68 132 238 17 156
Row24	107 121 161 10 161 9 20 198 169 138 48 106 117 66 154 212 39 138 182 170 100 205
Row25	239 134 4 241 162 117 207 144 82 110 159 225 223 146 117 77 58 91 177 12 235 170 169 25 58 91 198 126 163 134 109 159 163 42
Row26	184 55 118 129 109 36 170 237 189 210 206 187 92 161 1 1 65 173 93 233 3 70 109 215 74 126 112 221 209 13 52 66 106 154 80 175 188 225 60 22...
Row27	240 92 94 191 37 136 124 194 43 200 198 130 117 240 42 136 50 110 154 12 235 147 63 163 77 122 70 105 229 223 63 150 101 196 84 204 4 16 129...
Row28	159 155 33 163 1 23 161 216 225 130 191 92 205 141 60 215 145 12 235 47 117 226 1 170 207 176 128 41 188 99 79
Row29	52 60 161 138 48 74 84 60 60 20 209 158 41 104 48 62 225
Row30	92 200 230 63 77 229 231 87 107 241 33 12 235 93 229 69 58 91 84 211 173 103 31 68 104 184 70 158 133 216 92 17 227 184 117 212 242 215 234 ...
Row31	48 232 195 92 134 229 108 162 94 173 144 117 161 209 227 11 85 69 27 83 193 11 207
Row32	121 170 17 84 241 123 197 117 2 201 0 106 107 225 156 129 88 159 50 104 48 55 132 149 83 234 239 173 240 173 207 84 107 188 62 30 4
Row33	12 235 231 117 242 229 76 28 120 109 237 70 134 128 93 173 210 62 231 93 219 79 178 158 12 235 203 88 24 117 37 121
Row34	112 196 211 32 144 84 178 188 134 157 216 118 228 48 70
Row35	209 79 107 203 34 84 17 50 48 17 17 146 221 107 128 151 211 68 237 199 75
Row36	62 154 226 226 173 42 11 60 118 207 80 163 173 198 147 130 109 93
Row37	177 184 32 102 163 110 180 118 82 234 4 85 185 2 184 190 211 44 118 104 93 99 144 52 69 10 191

After uploading the data set on KNIME, we used “CellSplitter” to make these data which was separated by blanks into different values them and combine them to make an array to apply algorithms through “CreateCollectionColumn”. As an one more step of preprocessing, we used “ColumnFilter” not to consider the first column of data set which looks like the column above. In this project, we used “AssociationRuleLearner” which learns itself frequent itemsets and association rules and “Apriori” algorithm to see the results of data analysis after all used preprocessing steps.



Item column: [...] AggregatedValues

Algortihm: ● Apriori ○ FPgrowth ○ RElim ○ SaM ○ JIM ○ DICE ○ TANIMOTO

Target Type: ● Frequent ○ Closed ○ Maximal

Item set settings

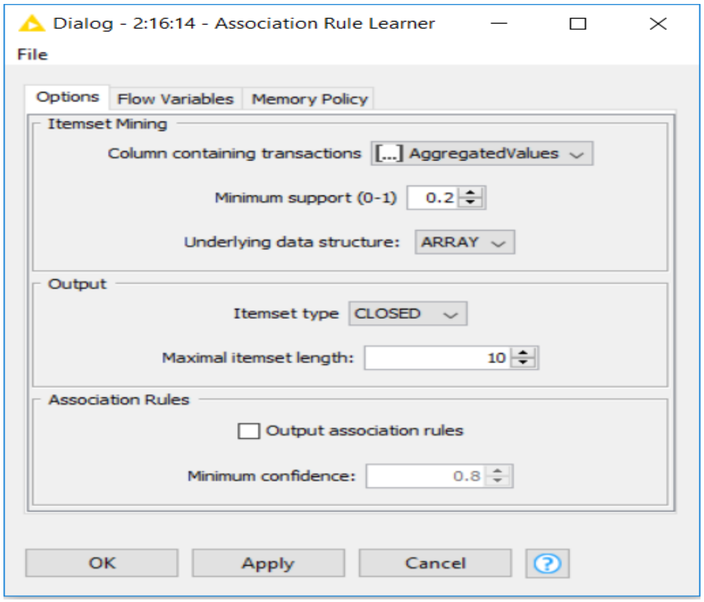
Minimum set size: 1

Minimum support: 5.0

○ Absolute number ● Percentage

Threshold: (optional) 10.0

□ Sort item set



Itemset Mining

Column containing transactions [...] AggregatedValues

Minimum support (0-1) 0.2

Underlying data structure: ARRAY

Output

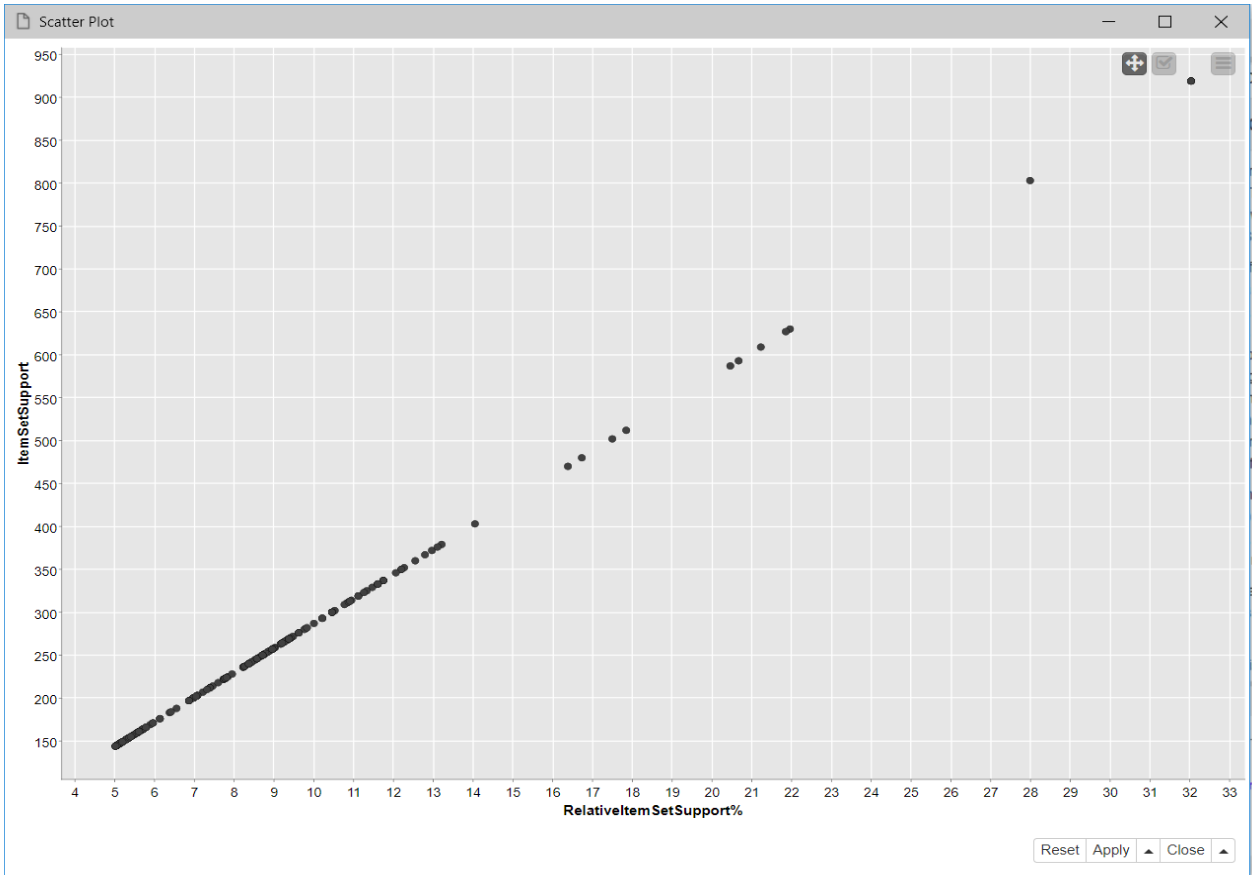
Itemset type CLOSED

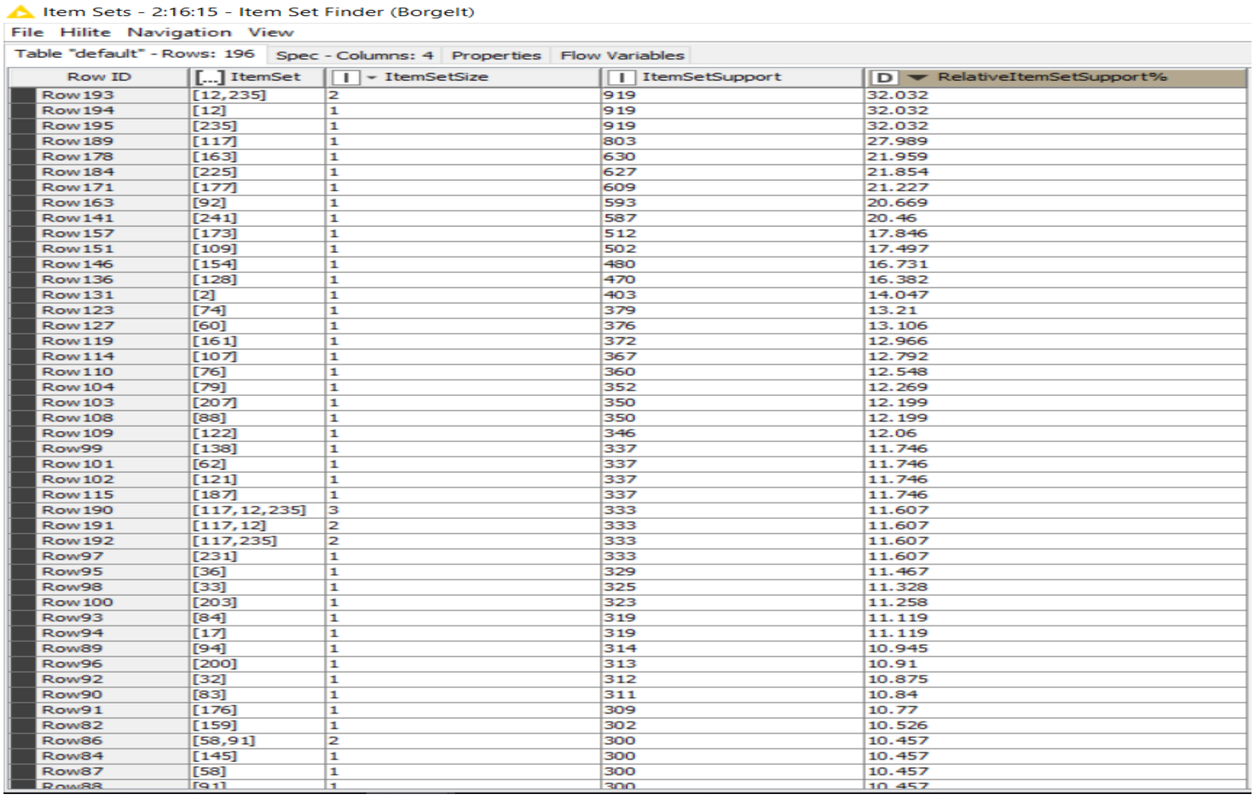
Maximal itemset length: 10

Association Rules

□ Output association rules

Minimum confidence: 0.8





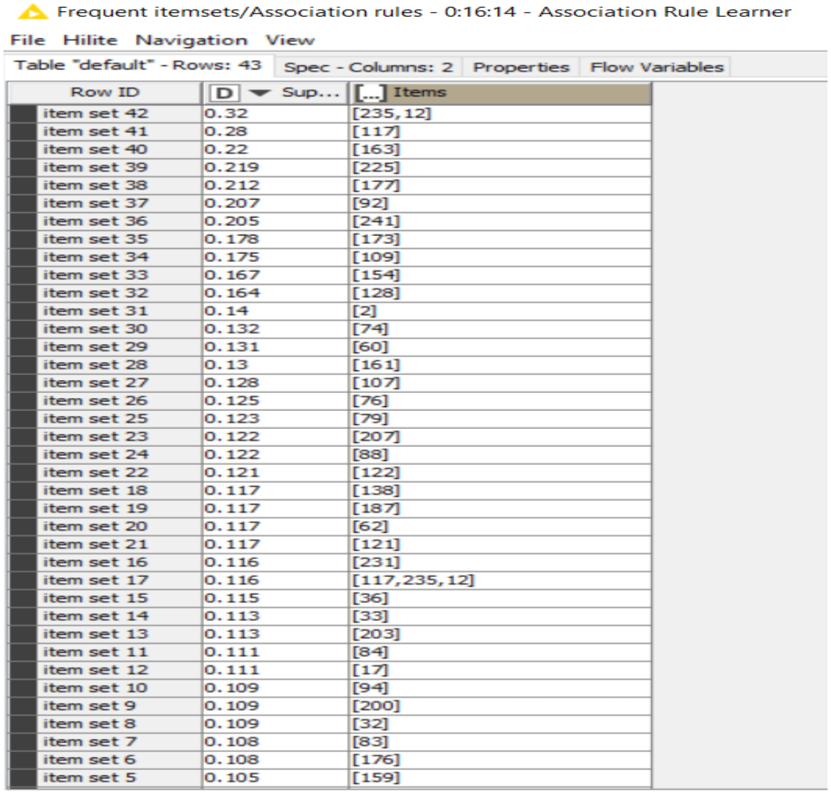
Item Sets - 2:16:15 - Item Set Finder (Borgelt)

File Hilite Navigation View

Table "default" - Rows: 196 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	[...] ItemSet	I ItemSetSize	I ItemSetSupport	D RelativeItemSetSupport%
Row193	[12,235]	2	919	32.032
Row194	[12]	1	919	32.032
Row195	[235]	1	919	32.032
Row189	[117]	1	803	27.989
Row178	[163]	1	630	21.959
Row184	[225]	1	627	21.854
Row171	[177]	1	609	21.227
Row163	[92]	1	593	20.669
Row141	[241]	1	587	20.46
Row157	[173]	1	512	17.846
Row151	[109]	1	502	17.497
Row146	[154]	1	480	16.731
Row136	[128]	1	470	16.382
Row131	[2]	1	403	14.047
Row123	[74]	1	379	13.21
Row127	[60]	1	376	13.106
Row119	[161]	1	372	12.966
Row114	[107]	1	367	12.792
Row110	[76]	1	360	12.548
Row104	[79]	1	352	12.269
Row103	[207]	1	350	12.199
Row108	[88]	1	350	12.199
Row109	[122]	1	346	12.06
Row99	[138]	1	337	11.746
Row101	[62]	1	337	11.746
Row102	[121]	1	337	11.746
Row115	[187]	1	337	11.746
Row190	[117,12,235]	3	333	11.607
Row191	[117,12]	2	333	11.607
Row192	[117,235]	2	333	11.607
Row97	[231]	1	333	11.607
Row95	[36]	1	329	11.467
Row98	[33]	1	325	11.328
Row100	[203]	1	323	11.258
Row93	[84]	1	319	11.119
Row94	[17]	1	319	11.119
Row89	[94]	1	314	10.945
Row96	[200]	1	313	10.91
Row92	[32]	1	312	10.875
Row90	[83]	1	311	10.84
Row91	[176]	1	309	10.77
Row82	[159]	1	302	10.526
Row86	[58,91]	2	300	10.457
Row84	[145]	1	300	10.457
Row87	[58]	1	300	10.457
Row88	[91]	1	300	10.457

The results of data analysis above shows the results of Apriori algorithm. If we look "ItemSet" column,we see that some of item sets which are greater than 1 have high "RelativeItemSetSupport". It shows that if we sell items [12,235] together,32.032 percent of customers buy that or if we sell items [117,12,235] together, 11.607 percent of customers buy. If we check the results from "Association Rule Learner", we get the same results after applying Apriori algorithm to the data set.



Frequent itemsets/Association rules - 0:16:14 - Association Rule Learner

File Hilite Navigation View

Table "default" - Rows: 43 | Spec - Columns: 2 | Properties | Flow Variables

Row ID	D Sup...	[...] Items
item set 42	0.32	[235,12]
item set 41	0.28	[117]
item set 40	0.22	[163]
item set 39	0.219	[225]
item set 38	0.212	[177]
item set 37	0.207	[92]
item set 36	0.205	[241]
item set 35	0.178	[173]
item set 34	0.175	[109]
item set 33	0.167	[154]
item set 32	0.164	[128]
item set 31	0.14	[2]
item set 30	0.132	[74]
item set 29	0.131	[60]
item set 28	0.13	[161]
item set 27	0.128	[107]
item set 26	0.125	[76]
item set 25	0.123	[79]
item set 23	0.122	[207]
item set 24	0.122	[88]
item set 22	0.121	[122]
item set 18	0.117	[138]
item set 19	0.117	[187]
item set 20	0.117	[62]
item set 21	0.117	[121]
item set 16	0.116	[231]
item set 17	0.116	[117,235,12]
item set 15	0.115	[36]
item set 14	0.113	[33]
item set 13	0.113	[203]
item set 11	0.111	[84]
item set 12	0.111	[17]
item set 10	0.109	[94]
item set 9	0.109	[200]
item set 8	0.109	[32]
item set 7	0.108	[83]
item set 6	0.108	[176]
item set 5	0.105	[159]